10/522675
JT12 Rec'd PCT/PTO 27 JAN 2005

Site-specific Labelling of Proteins using Cyanine Dye Reporters

The present invention relates to reagents and methods for site-specific labelling of proteins using cyanine dyes as reporter molecules. In particular, the invention relates to new cyanine dye derivatives containing thioester activated groups and groups reactive with target molecules containing or derivatised to contain a thioester reactive moiety.

There is increasing interest in, and demand for, fluorescent reporters for use in the labelling and detection of biomolecules. Cyanine and related dyes such as rigidised cyanine dyes and squaraines offer a number of advantages over other fluorescent dye reagents and they are finding widespread use as fluorescent labels in such diverse areas as sequencing, microarrays, flow cytometry and proteomics. For example, US 5569587 (Waggoner et al) discloses water soluble cyanine dye derivatives that possess reactive groups suitable for reaction with target molecules that contain, or are derivatised to contain, –OH, $-NH_2$, or –SH groups. The cyanine dyes are characterised by having very high extinction coefficients and favourable quantum yields. In addition, cyanine dyes possess good photostability and are not readily photobleached.

In many applications there is a need to form a permanent link, in the form of a covalent bond, between a fluorescent labelling dye and a target molecule such as a protein. The chemistry of peptide and protein labelling is well documented and a wide range of labelling reagents are now commercially available. For a review and examples of protein labelling using fluorescent labelling reagents, see "Non-Radioactive Labelling, a Practical Introduction", Garman, A.J. Academic Press, 1997; "Handbook of Fluorescent Probes and Research Chemicals", Haugland, R.P., Molecular Probes Inc., 1992).

Site-specific incorporation of a fluorescent label into a protein or peptide may be of considerable benefit in certain biochemical and biophysical studies, for example fluorescence resonance energy transfer, and protein

structure and function studies. One method for the site-specific attachment of a fluorescent label into a target polypeptide utilises the native chemical ligation reaction. According to this procedure, an unprotected peptide fragment containing an N-terminal cysteine residue and a second unprotected peptide fragment containing an α-thioester group are chemoselectively ligated together at physiological pH, irrespective of their primary sequences, to generate an amide bond at the ligation site. For examples, see reviews by Cotton, G.J. and Muir T.W., Chem. Biol., (1999), 6, R247-260; Giriat, I., Muir, T.W. and Perler, F.B., Genetic Engineering, (2001), 23, 171-199; Muir, T.W., Syn. Lett., (2001), 6, 733-740.

Tolbert, T.J. and Wong, C-H. (Angew. Chem. Int. Ed., (2002), 41, 2171-2174) describe the preparation of fluorescein and biotin thioester derivatives and the reaction of these with N-terminal cysteine-containing recombinant proteins. Schuler, B. and Pannell, L.K. (Bioconjugate Chemistry, published on line, 18 July 2002) reported the preparation of a benzyl thioester of Cy5™ and subsequent reaction with a synthetic polypeptide containing an N-terminal cysteine residue.

However, there are no reports describing thioester derivatives of cyanine dyes in which the reporter is also linked covalently to an affinity tag. Use of such a reagent in reactions involving site specific labelling of proteins and peptides will be advantageous for subsequent separation and purification of the fluorescent dye-labelled target. The present invention therefore provides new cyanine dye reagents and methods that afford direct attachment of the cyanine dye reporter to either the N-terminus or C-terminus of a synthetic or recombinant peptide or protein and their derivatives, in a site-specific manner, coupled with purification of the resultant labelled molecule.

According to one aspect of the present invention, there is provided a compound comprising a cyanine dye or derivative thereof containing at least one target bonding group selected from a carboxylic acid thioester group or a

group suitable for covalent reaction with a thioester, characterised in that said compound includes an affinity tag covalently bound thereto.

Suitably, the compound is of formula (I):

$$\begin{array}{c} D - L^{1} - M - L^{2} - B \\ \quad\;\; | \\ \quad\;\; F \end{array}$$

(I)

wherein:

D is a dye selected from a cyanine dye or a derivative thereof;

B is an affinity tag;

F comprises a target bonding group selected from a carboxylic acid thioester group and a 1,2-aminothiol group;

M is a group adapted for attaching to F; and

L^1 and L^2 each independently comprise a group containing from 1 – 40 linked atoms selected from carbon atoms which may optionally include one or more groups selected from –NR'–, –O–, –CH=CH–, –CO–NH– and phenylenyl groups, where R' is selected from hydrogen and C_1 – C_4 alkyl.

Suitably, there are 2 to 30 atoms in each of L^1 and L^2, preferably, 6 to 20 atoms.

Preferably, L^1 and L^2 are independently selected from the group:

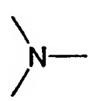
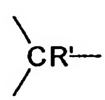
$$-\{(CHR')_p-Q-(CHR')_r\}_s-$$

where Q is selected from: –CHR'–, –NR'–, –O–, –CH=CH–, –Ar– and –CO–NH–; R' is hydrogen or C_1 – C_4 alkyl, p is 0 – 5, r is 1 – 5 and s is 1 or 2.

Particularly preferred Q is selected from: –CHR'–, –O– and –CO–NH–, where R' is hereinbefore defined.

In one embodiment L^2 is a cleavable linker and may additionally include group P which may be suitably selected from a chemically-cleavable group, an enzyme-cleavable group, or a photochemically-cleavable group. Suitable chemically cleavable groups include carbamate esters and carboxylate esters, which are both cleaved under basic conditions. Suitable enzyme cleavable groups may be selected from groups such as ester, amide and phospho-diester groups. Such groups are substrates for, and are hydrolysed by hydrolases, such as proteases, esterases and phospho-diesterases. Suitable photocleavable groups P for use in the compound of formula (I) may contain the 4,5-dialkoxy-2-nitrobenzyl alcohol linker (Holmes, C.P., and Jones, D.G., J.Org.Chem., (1995), 60, 2318-2319) or phenacyl linkers (Wang, S., J.Org.Chem., (1976), 41, 3258-3261). These groups undergo efficient photoreaction upon 300nm illumination, resulting in the rapid cleavage of the dye molecule or dye-labelled protein from the affinity tag.

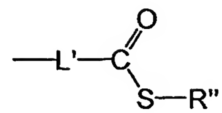
Suitably, the group M may be any suitable functional group adapted for attaching the target bonding group F. Preferably, M is selected from:

 and 

wherein R' is hereinbefore defined.

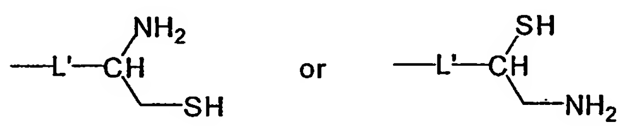
Suitable affinity tags may be selected from biotin, desthiobiotin and metal chelating ligands such as his-tag and iminodiacetic acid, nitrilotriacetic acid and the like. Preferred affinity tags are selected from biotin and desthiobiotin.

In one embodiment of the present invention, the target bonding group F is a carboxylic acid thioester of formula:



wherein L' is a bond or is a group containing from 1 – 30 linked atoms selected from carbon atoms and optionally one or more groups selected from –NH–, –O– and –CO–NH–; and R" is $C_1 - C_4$ alkyl, $C_6 - C_{10}$ aryl, or $C_7 - C_{15}$ aralkyl, which may be optionally substituted with sulphonate; or is the group $-(CH_2)_2-CONH_2$. In the case where L' is a bond, the target bonding group F is attached directly to group M.

In an alternative embodiment, the target bonding group F is a 1,2-aminothiol group of formula:



wherein L' is hereinbefore defined.

Thus, the present invention provides fluorescent labelling reagents comprising a cyanine dye or derivative thereof, that are modified by incorporating a target bonding group and an affinity tag into the molecule. The target bonding group may be selected from a carboxylic acid thioester group or a 1,2-aminothiol group. Where the target bonding group is a thioester group, it is selectively reactive with a 1,2-aminothiol group on a target molecule, suitably a protein or peptide, or a derivative thereof. In the alternative, the cyanine dye may contain a 1,2-aminothiol group for reaction with a thioester group on the target. The incorporation of a reactive thioester or, alternatively, a 1,2-aminothiol functionality into the chemical structure of the reporter molecule enables the target molecule to be directly labelled in a convenient one step process. According to the methods of the invention, labelling of peptides and proteins is site-specific, irrespective of the composition of the primary sequence. By generating the target primary

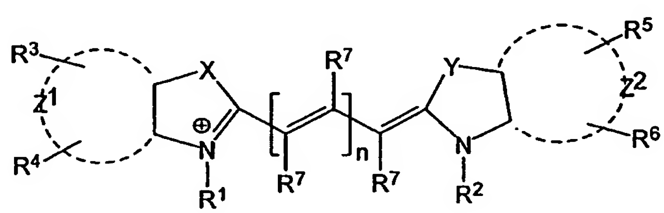
sequence with either an N-terminal cysteine or a thioester functionality, site-specific labelling can be achieved directly, by incubating the target with the appropriate derivative of the cyanine dye, suitably, the thioester and 1,2-aminothiol derivatives respectively. Furthermore, inclusion of an affinity tag in the labelling reagent allows subsequent purification of the fluorescent dye labelled protein or peptide.

Suitably, the cyanine dye or cyanine dye derivative may be selected from cyanine dyes, rigidised cyanine dyes and squaraine dyes, provided that the dye incorporates at least one carboxylic acid thioester group, or a group suitable for covalent reaction with a thioester. Table 1 shows some examples of cyanine dyes, having particular excitation (Abs) and emission (Em) characteristics.

Table 1

Dye	Fluorescence Colour	Abs (nm)	Em (nm)
Cy2	Green	489	506
Cy3	Orange	550	570
Cy3.5	Scarlet	581	596
Cy5	Far red	649	670
Cy5.5	Near-IR	675	694
Cy7	Near-IR	743	767

In one embodiment according to the first aspect, the compound has the formula (II):



(II)

wherein:

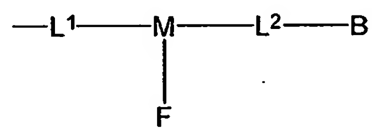
groups R^3 and R^4 are attached to the Z^1 ring structure and groups R^5 and R^6 are attached to the Z^2 ring structure;

n is an integer from 1 to 3;

Z^1 and Z^2 independently represent the atoms necessary to complete one ring or two fused ring aromatic or heteroaromatic systems, each ring having five or six atoms selected from carbon atoms and optionally no more than two atoms selected from oxygen, nitrogen and sulphur;

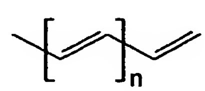
X and Y are the same or different and are selected from: $>CR^8R^9$, oxygen, sulphur, –CH=CH–, >N–W wherein N is nitrogen and W is selected from hydrogen and the group R^{10};

at least one of groups R^1, R^2, R^3, R^4, R^5, R^6, R^8, R^9 and R^{10} is the group:



where B, F, M, L^1 and L^2 are hereinbefore defined;

groups R^7 are independently selected from hydrogen and $C_1 - C_4$ alkyl which may be unsubstituted or substituted with aryl, or two or more of R^7 together with the group:



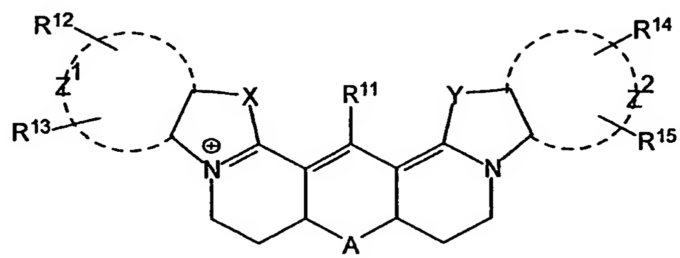
form a hydrocarbon ring system substituted with R^7 and which may optionally contain a heteroatom selected from –O–, –S– or $>NR^7$, wherein R^7 and n are hereinbefore defined;

remaining groups R^3, R^4, R^5 and R^6 are independently selected from the group consisting of hydrogen, halogen, amide, cyano, nitro, mono- or di-$C_1 - C_6$ alkyl-substituted amino, carbonyl, carboxyl, $C_1 - C_6$ alkyl, $C_1 - C_6$ alkoxy, aryl, heteroaryl, aralkyl and the group $-(CH_2)_m-Y$ where Y is selected from sulphonate, sulphate, phosphonate, phosphate and quaternary ammonium and m is zero or an integer from 1 to 6;

remaining groups R^8, R^9 and R^{10} are independently $C_1 - C_6$ alkyl; and

remaining groups R^1 and R^2 are independently selected from hydrogen, C_1 – C_{10} alkyl, the group $-(CH_2)_m-Y$ wherein Y and m are hereinbefore defined, and benzyl which may be unsubstituted or substituted by up to two nitro groups.

In a second embodiment according to the first aspect, the compound has the formula (III):



(III)

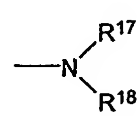
wherein

groups R^{12}, R^{13}, R^{14} and R^{15} are attached to the rings containing X and Y or, optionally are attached to atoms of the Z^1 and Z^2 ring structures;

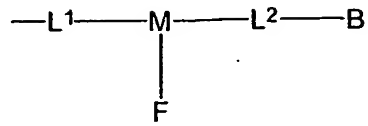
Z^1 and Z^2 independently represent the atoms necessary to complete one ring or two fused ring aromatic or heteroaromatic systems, each ring having five or six atoms selected from carbon atoms and optionally no more than two atoms selected from oxygen, nitrogen and sulphur;

X and Y are the same or different and are selected from: $>CR^8R^9$, oxygen, sulphur, –CH=CH–, >N–W wherein N is nitrogen and W is selected from hydrogen and the group R^{10};

A is selected from O and NR^{16} where R^{16} is the substituted amino radical:



at least one of groups R^8, R^9, R^{10} R^{11}, R^{12}, R^{13}, R^{14}, R^{15}, R^{17} and R^{18} is the group:



where B, F, M, L^1 and L^2 are hereinbefore defined;
remaining groups R^{11}, R^{12}, R^{13}, R^{14} and R^{15} are independently selected from the group consisting of hydrogen, halogen, amide, cyano, nitro, amino, mono- or di-$C_1 - C_6$ alkyl-substituted amino, carbonyl, carboxyl, $C_1 - C_6$ alkyl, $C_1 - C_6$ alkoxy, aryl, heteroaryl, aralkyl and the group $-(CH_2)_m-Y$ where Y is selected from sulphonate, sulphate, phosphonate, phosphate and quaternary ammonium and m is zero or an integer from 1 to 6;
remaining groups R^8, R^9 and R^{10} are independently $C_1 - C_6$ alkyl;
remaining group R^{17} is selected from hydrogen, $C_1 - C_4$ alkyl and aryl; and
remaining group R^{18} is selected from $C_1 - C_6$ alkyl, aryl, heteroaryl, an acyl radical having from 2-7 carbon atoms, and a thiocarbamoyl radical.

Suitably, in the compounds according to formula (II) and (III), Z^1 and Z^2 may be selected independently from the group consisting of phenyl, pyridinyl, naphthyl, anthranyl, indenyl, fluorenyl, quinolinyl, indolyl, benzothiophenyl, benzofuranyl and benzimidazolyl moieties. Additional one, or two fused ring systems will be readily apparent to the skilled person. Preferably, Z^1 and Z^2 are selected from the group consisting of phenyl, pyridinyl, naphthyl, quinolinyl and indolyl moieties. Particularly preferred Z^1 and Z^2 are phenyl and naphthyl moieties.

Suitably, at least one of the groups R of the compounds of formula (II) and (III) is a water solubilising group for conferring a hydrophilic characteristic to the compound. Solubilising groups, for example, sulphonate, sulphonic acid and quaternary ammonium, may be attached directly to the aromatic ring structures Z^1 and/or Z^2 of the compounds of formula (II) and (III). Alternatively, solubilising groups may be attached by means of a C_1 to C_6 alkyl linker chain to said aromatic ring structures and may be selected from the group $-(CH_2)_m-Y$ where Y is selected from sulphonate, sulphate,

phosphonate, phosphate, quaternary ammonium and carboxyl; and m is hereinbefore defined. Alternative solubilising groups may be carbohydrate residues, for example, monosaccharides, or polyethylene glycol derivatives. Examples of water solubilising constituents include $C_1 - C_6$ alkyl sulphonates, such as $-(CH_2)_3-SO_3^-$ and $-(CH_2)_4-SO_3^-$. However, one or more sulphonate or sulphonic acid groups attached directly to the aromatic ring structures of a dye of formula (II) or (III) are particularly preferred. Water solubility may be advantageous when labelling proteins.

In one embodiment the compound of formula (I) is a fluorescent reporter molecule. In this embodiment, none of the substituent groups R in the compounds of formula (II) and (III) contains a nitro group.

In another embodiment, the compound of formula (I) is non-fluorescent or substantially non-fluorescent dye wherein at least one of the groups R attached to the aromatic ring structures of the compounds of formula (II) and (III) comprises at least one nitro group. In this embodiment, suitably, the at least one nitro group may be attached directly to the Z^1 and/or Z^2 ring structures. In the alternative, a mono- or di-nitro-substituted benzyl group may be attached to the Z^1 and/or Z^2 ring structures which optionally may be further substituted with one or more nitro groups. The non-fluorescent or substantially non-fluorescent cyanine dye or cyanine dye derivatives according to the invention may be used to label one component of a fluorescent donor/acceptor pair in assays involving the detection of binding and/or cleavage events in reactions involving biological molecules, as described in EP 1086179 B1 (Amersham Biosciences UK Limited).

In the embodiments according to the first aspect:

i) Aryl is an aromatic substituent containing one or two fused aromatic rings containing 6 to 10 carbon atoms, for example phenyl or naphthyl, the aryl being optionally and independently substituted by one or more substituents, for example halogen, straight or branched chain alkyl groups

containing 1 to 10 carbon atoms, aralkyl and alkoxy for example methoxy, ethoxy, propoxy and n-butoxy;

ii) Heteroaryl is a mono- or bicyclic 5 to10 membered aromatic ring system containing at least one and no more than 3 heteroatoms which may be selected from N, O, and S and is optionally and independently substituted by one or more substituents, for example halogen, straight or branched chain alkyl groups containing 1 to 10 carbon atoms, aralkyl and alkoxy for example methoxy, ethoxy, propoxy and n-butoxy;

iii) Aralkyl is a $C_1 - C_6$ alkyl group substituted by an aryl or heteroaryl group;

iv) Halogen and halo groups are selected from fluorine, chlorine, bromine and iodine.

By virtue of the target bonding group F, the compounds according to the present invention are useful for covalently labelling target biological materials in a site specific manner for applications in biological detection systems. Suitable target materials include proteins, post-translationally modified proteins, peptides, antibodies, antigens, and protein-nucleic acids (PNAs). The reporter moiety may also be conjugated to species which can direct the path of the reporter within or aid entry to or exit from cells (live or dead); such as for example, long alkyl residues to allow permeation of lipophilic membranes, or intercalating species to localise a reporter in a nucleus or other cellular enclave containing double-stranded DNA.

In a second aspect, there is provided a method for labelling a protein of interest wherein said protein contains or is derivatised to contain an N-terminal cysteine, the method comprising:

i) adding to a liquid containing said protein a compound of formula (I):

$$\begin{array}{c} D - L^1 - \underset{\displaystyle F}{\underset{|}{M}} - L^2 - B \end{array}$$

(I)

wherein:

D is a dye selected from a cyanine dye or a derivative thereof;

B is an affinity tag;

F comprises a target bonding group selected from a carboxylic acid thioester group and a 1,2-aminothiol group;

M is a group adapted for attaching to F; and

L^1 and L^2 each independently comprise a group containing from 1 – 40 linked atoms selected from carbon atoms which may optionally include one or more groups selected from –NR'–, –O–, –CH=CH–, –CO–NH– and phenylenyl groups, where R' is selected from hydrogen and C_1 – C_4 alkyl; and

ii) incubating said compound with said protein under conditions suitable for labelling said protein.

Suitably, there are 2 to 30 atoms in each of L^1 and L^2, preferably, 6 to 20 atoms.

Preferably, L^1 and L^2 are independently selected from the group:

$$-\{(CHR')_p-Q-(CHR')_r\}_s-$$

where Q is selected from: –CHR'–, –NR'–, –O–, –CH=CH–, –Ar– and –CO–NH–; R' is hydrogen or C_1 – C_4 alkyl, p is 0 – 5, r is 1 – 5 and s is 1 or 2.

Particularly preferred Q is selected from: –CHR'–, –O– and –CO–NH–, where R' is hereinbefore defined.

Preferred compounds of formula (I) for use in labelling a target protein are those having formula (II) or (III) as hereinbefore defined.

Covalent labelling using compounds of the present invention may be accomplished with a target having at least one carboxylic acid thioester group or 1,2-aminothiol group as hereinbefore defined. The target may be incubated

with an amount of a compound of the present invention having at least one group F as hereinbefore defined that can covalently bind with the complementary group of the target material. The target material and the compound of the present invention are incubated under conditions and for a period of time sufficient to permit the target material to covalently bond to the compound of the present invention. Thus, for example, the thioester group F may be reacted and form a covalent bond with any of the above target materials that contains, or has been derivatised to contain, a 1,2-amino thiol group. These methods and the products resulting from them, for example, reporter-labelled biomolecules are envisaged as further aspects of the invention.

Suitably, the protein of interest may be selected from the group consisting of antibody, antigen, protein, peptide, microbial materials, cells and cell membranes.

In a particular embodiment according to the second aspect, there is provided a method of separating and/or purifying the dye-labelled protein of interest by affinity chromatography utilising the affinity of the affinity tag moiety for an immobilised ligand (or specific binding partner) attached to a support material. Affinity chromatography provides a quick and convenient method to enable the separation of labelled and unlabelled protein molecules under physiological conditions. Proteins labelled with an affinity tag can be selectively bound to an affinity column and any unreacted protein removed by washing the column. Suitable specific binding moieties include avidin or streptavidin (for a biotin tag); immobilised metal ions, for example, Cu(II), Ni(II), Fe(II) and Fe(III) (for His-tag or iminodiacetic acid). Methods for affinity purification of proteins will be well known to the skilled person, see for example Ostrove, S, Methods in Enzymology, (1990), Vol 182, page 357.

In a typical labelling procedure, a target peptide or protein containing an N-terminal cysteine residue is agitated with an excess of a cyanine dye thioester derivative, e.g. Cy5-MESNA (Cy5-mercaptoethanesulphonic acid

ester), in phosphate buffer (typically 200 mM NaCl, 200 mM sodium phosphate) at ~pH 7.3 – 7.4 containing ~1.5% MESNA. The concentration of the target polypeptide in the labelling reaction is generally between 100 μM to 10 mM, whilst the Cy5-MESNA is generally present in excess, for example 1.5 to 3-fold molar excess. When the target polypeptide concentration is relatively low the concentration of Cy5-MESNA is usually maintained at or above 1 mM. Generally, for labelling small peptides a solution of Cy5-MESNA and MESNA cofactor is directly added to the lyophilised target.

Typically, for site specific labelling of proteins and large polypeptides using the reagents of the present invention, the target is first exchanged into an appropriate buffer, which is known not to affect the labelling reaction. An equal volume of a solution of Cy5-MESNA and MESNA thiol cofactor in ligation buffer is then added to the protein to give the desired final concentration of the reactants. The reaction mixture is agitated overnight at room temperature. The reaction time may be lowered to less than one hour for high reactant concentrations or, if the stability of the target polypeptide is an issue, the labelling reaction may be performed efficiently at 4°C. On completion of the labelling reaction, dithiothreitol (DTT) is added to a final concentration of ~50 mM and the desired material isolated by affinity chromatography.

Various different denaturants, organic solvents and detergents may be added to the reaction buffer when performing native chemical ligation and expressed protein ligation reactions, to aid the ligation of the peptide fragments and/or stabilise the reactants or products. Such reagents may be utilised in the labelling reaction to increase product yield if necessary. Examples include, but are not limited to guanidinium chloride, urea, dimethylformamide, dimethylsulphoxide, acetonitrile, triton X-100, octyl glucoside, 1,6-hexanediol and glycerol.

The ligation reaction using the derivatised cyanine dye according to the present invention may be optimally performed at between pH 7.0 and pH 8.0

and at temperatures varying between 4°C and 37°C. It is envisaged that such a range of conditions are compatible to the site-specific labelling reaction described herein.

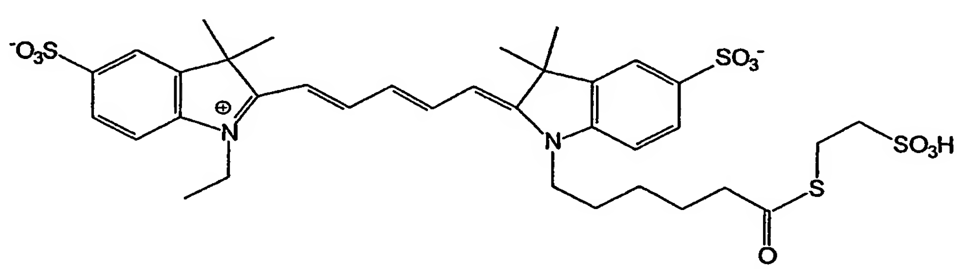
The advantage of the present method is that it enables the introduction of an extrinsic label into a proteinacious substrate in a regioselective and specific manner, thus minimising any detrimental effects that labelling may have on the biological function of the protein. The importance of controlling stoichiometry of labelling is important where dye overload may interfere with biological activity. In addition, if this controlled labelling stoichiometry is directed towards a single terminal site, rather than towards an internal site, this may have the benefit of further maintaining the biological viability of the labelled species.

The invention is further illustrated by reference to the following examples and figure in which:

Figure 1 illustrates the products from the labelling reaction of an N-terminal cysteine derivative of the Grb2SH2 domain with the thioester derivative, α-D-desthiobiotin-ε-Cy5-L-lysine-MESNA according to Examples 3 and 4.

<u>Experimental</u>

1. <u>2-[(1*E*,3*E*,5*E*)-5-(3,3-Dimethyl-1-{6-oxo-6-[(2-sulphoethyl)thio]hexyl}-5-sulfo-1,3-dihydro-2*H*-indol-2-ylidene)penta-1,3-dienyl]-1-ethyl-3,3-dimethyl-5-sulfo-3*H*-indolium</u>



To Cy™5 mono acid (47mg, 0.062mmol) in a solution of 7-azobenzotriazolyoxytris(pyrrolidino)phosphonium hexafluorophosphate (PyAOP, 66mg, 0.127mmol) in anhydrous dimethylformamide (DMF, 1ml) was added anhydrous di-isopropylethylamine (DIEA)(30µl, 0.1724mmol) and mixed for 5 minutes. The activated dye solution was then added to a stirred solution of 2-mercaptoethanesulphonic acid, sodium salt (MESNA, 40mg, 0.243mmol) in DMF (2mls) and DIEA (30µl, 0.1724mmol) under a dry nitrogen atmosphere. To this mixture was added as a solid, dried 4A molecular sieves(~1g, <5micron, activated powder). The mixture was stirred under a dry nitrogen atmosphere, at room temperature, in the dark overnight. Thin layer chromatography analysis (reverse phase C18 plates, eluents water/acetonitrile (70:30, containing 0.1% TFA) indicated a major component, $Rf_{thioester}$ = 0.25) with no trace of starting material (Rf_{acid} = 0.12).

The molecular sieves were removed by filtration and filtrate was added dropwise into an excess of ethyl acetate, the blue solid was filtered off and was purified by reverse phase-high performance liquid chromatography (RP-HPLC); [Phenomenex Prodigy C18 column; 15%B-30%B over 30 mins @ 20 ml/min; eluent A = 0.1%TFA/water, eluent B = 0.1%TFA.MeCN, UV detection at 650nm]. The product was isolated as a dark blue/purple solid (40 mg, 0.0513mmol, 83 % yield).

Accurate mono-isotopic mass: $C_{35}H_{45}O_{10}N_2S_4$ requires 781. Found Maldi Tof, LC-MS found mass: M+ 781.25. δ H (300MHz, d6-DMSO): 8.37 (t, 1H), 8.36 (t, 1H), 7.83 (d, 1H), 7.82 (d, 1H), 7.67(dd, 1H), 7.64 (dd, 1H), 7.36 (d, 1H), 7.33 (d, 1H), 6.61 (t, 1H), 6.38 (d, 1H), 6.28 (d, 1H), 4.15 (m, 2H), 4.08 (t, 2H), 3.06 (m, 2H), 2.63 (m, 2H), 2.56 (t, 2H), 1.64 (m, 2H), 1.28(t, 3H, 7.1), 1.40 (m, 2H). λ_{max} (abs) = 647nm. (ε (H_2O) = 230,000$M^{-1}cm^{-1}$).

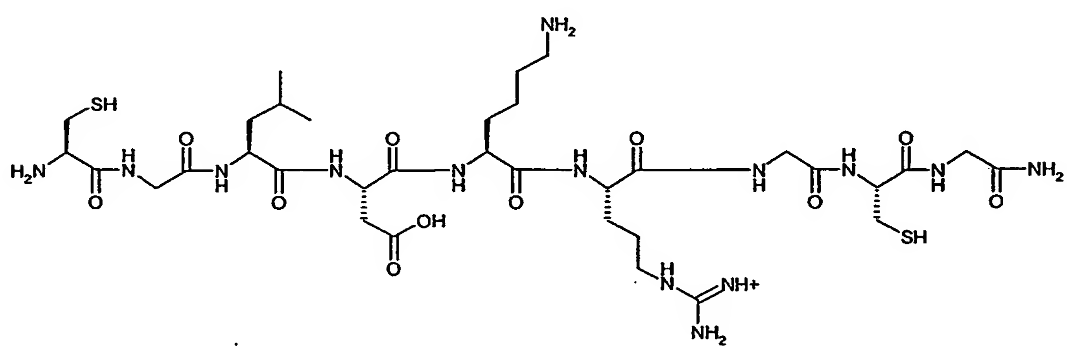
2. Determination of Specificity of Labelling using 2-[(1*E*,3*E*,5*E*)-5-(3,3-dimethyl-1-{6-oxo-6-[(2-sulphoethyl)thio]hexyl}-5-sulfo-1,3-dihydro-2*H*-indol-2-ylidene)penta-1,3-dienyl]-1-ethyl-3,3-dimethyl-5-sulfo-3*H*-indolium

2.1 Preparation of Cy5-Cys-Gly-Leu-Asp-Lys-Arg-Gly-Cys-Gly-NH_2

i) Synthesis of H-Cys(Trt)-Gly-Leu-Asp(OtBu)-Lys(Boc)-Arg(Pmc)-Gly-Cys(Trt)-Gly-rink amide resin

H-Cys(Trt)-Gly-Leu-Asp(OtBu)-Lys(Boc)-Arg(Pmc)-Gly-Cys(Trt)-Gly-rink amide resin was synthesised using a commercially available Applied Biosystems Model 433A automated peptide synthesiser using FastMocTM chemistry, following the instrument manufacturer's recommended procedures throughout. The peptide was synthesised on a 0.25 millimolar scale employing *O*-(benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) as the activating agent.

ii) H-Cys-Gly-Leu-Asp-Lys-Arg-Gly-Cys-Gly-NH_2



H-Cys(Trt)-Gly-Leu-Asp(OtBu)-Lys(Boc)-Arg(Pmc)-Gly-Cys(Trt)Gly-rink amide resin (100mg, theoretical loading 0.36mmol/g) was deprotected and cleaved from solid phase in 95% trifluoroacetic acid (TFA) / 2.5%tri-isopropylsilane (TIS) / 2.5% water (3 mls) at room temperature for 2 hours. The crude product was precipitated into a 10 fold excess of cold diethyl ether, centrifuged at 2500 rpm for 5 minutes and the ether decanted off. The crude

peptide was washed twice more with ether and was purified by reverse phase-high performance liquid chromatography (RP-HPLC) [Phenomenex Jupiter C18 column, eluent A: 0.1%TFA/water, eluent B: 0.1%TFA/acetonitrile, gradient : 0-73%B over 30 mins @1ml/min, detection at 214nm]. The product was isolated and lyophilised to afford a colourless fluffy solid (21mg by weight, 60%). Mono-isotopic mass: 906.4. Found mass (LC-MS): MH+ @ 907.3; M+Na @ 929.6; > 95% pure as judged by RP-HPLC @ 214nm (Phenomenex Jupiter C18 column, eluent A: 0.1%TFA/water, eluent B: 0.1%TFA/acetontrile, 5-50% B over 25mins @ 1ml /min, UV detection at 650nm).

iii) Cy5-Cys-Gly-Leu-Asp-Lys-Arg-Gly-Cys-Gly-NH_2

To solid H-Cys-Gly-Leu-Asp-Lys-Arg-Gly-Cys-Gly-NH2 (3.0mg by weight, 0.0033mmol) was added a solution of Cy5-MESNA (3.5mg, 0.0045mmol) in 200mM phosphate buffer, 200mM NaCl pH 7.2 containing 1.5% 2-mercaptoethanesulphonic acid, sodium salt (400μl). The reaction mixture was stirred on rollers for 30 minutes at room temperature in darkness. During incubation, a blue precipitate formed, which re-dissolved on addition of acetonitrile (40μl).

500mM DTT (200μl) in 200mM phosphate buffer; 200mM NaCl pH 7.2 (0.5mls, 0.0025mmol) was then added to the reaction mixture, with complete mixing and was stirred for a further 30 minutes at room temperature in the dark. The crude reaction mixture was then purified by RP-HPLC [Phenomenex Jupiter C18 column, eluent A: 0.1%TFA/water, eluent B: 0.1%TFA/acetontrile, gradient; 20-35%B over 30 mins at 4 ml/min, detection at 650nm and 214nm]. The product was isolated and lyophilised as a blue fluffy solid (1.6 mg by UV/VIS at 650nm; 50% Yield; 98% pure as judged by RP-HPLC at 650nm. Mono-isotopic mass $C_{67}H_{101}N_{16}O_{18}S_4$ requires 1545.636. Found (LC-MS) M+ 1545.7.

2.2 Characterisation of Labelled Peptide

i) Ellman's Test on Cy5-Cys-Gly-Leu-Asp-Lys-Arg-Gly-Cys-Gly-NH_2

A sample of Cy5-Cys-Gly-Leu-Asp-Lys-Arg-Gly-Cys-Gly-NH_2 was dissolved in 100mM sodium phosphate buffer; 1mM EDTA pH 7.27 (stock buffer) to afford a 0.3μM peptide stock by UV/VIS at 650nm.

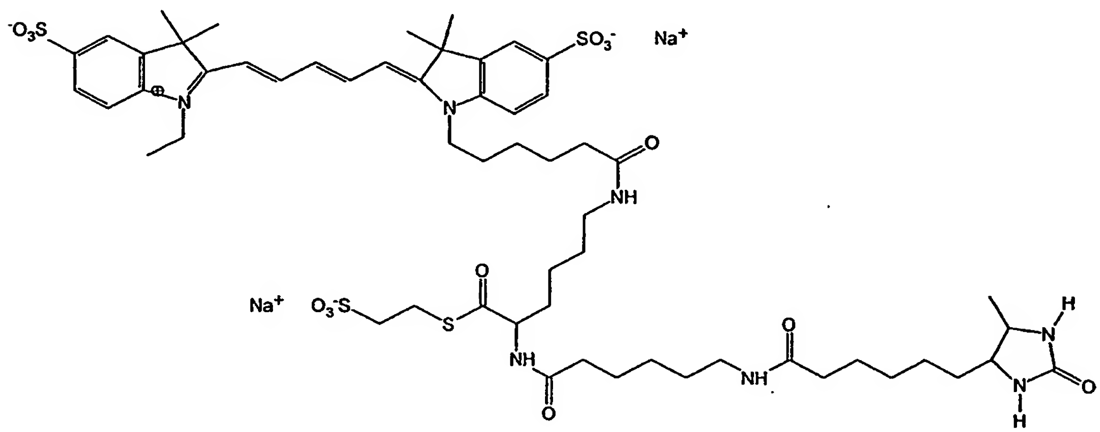
0.3μM peptide stock (40μl) and 10mM 5,5'-dithiobis(2-nitrobenzoic acid (DTNB) in 100mM sodium phosphate buffer; 1mM EDTA pH 7.27 (50μl) were mixed together in stock buffer (910μl) to afford a green solution. The absorbance at 412nm (due to generation of TNB^{2-}) was recorded against a DTNB blank [10mM DTNB stock (50μl) in stock buffer (950μl]. Using the known molar absorption coefficient of TNB^{2-} (14150$M^{-1}cm^{-1}$), the thiol concentration was determined as 655μM, approximately twice the peptide concentration, confirming two free thiol groups. [SH] = [A412nm (sample)-A412nm(reference)/ε (TNB^{2-})

ii) Enzyme Digestion of Cy5-Cys-Gly-Leu-Asp-Lys-Arg -Gly-Cys-Gly-NH_2

To a solution of Cy5-Cys-Gly-Leu-Asp-Lys-Arg-Gly-Cys-Gly-NH_2 (180μg by UV/VIS at 650nm) in TRIS buffer pH 8.0 (100μl) containing 10% acetonitrile was added Asp-N (2μg) in TRIS buffer pH 8.0 (70μl). The reaction mixture was stirred at room temperature in the dark for 4 hours. The reaction mixture was treated with 250mM Tris (2-carboxyethyl)phosphine, HCL (TCEP) in TRIS buffer pH 8.0 (55μl) for 30 minutes. The reduced reaction mixture was then diluted 1:5 with 0.1%TFA in water and purified by reverse phase HPLC [Phenomenex Jupiter C18, eluent A: 0.1%TFA/water, eluent B: 0.1%TFA/acetonitrile, 5-50% B over 30mins @ 1ml /min, UV at 214nm, 650nm]. The two components of the reaction mixture were identified as : Cy5-Cys-Gly-Leu-OH, mono-isotopic mass: $C_{44}H_{60}N_5O_{11}S_3$ requires 930.3451. Found mass (MALDI Tof): M+ 930.0 and H-Asp- Lys-Arg-Gly-Cys-Gly-NH_2,

monoisotopic mass: $C_{23}H_{43}N_{11}O_8S$ requires 633.3016, found mass (MALDI-Tof): M+ 633.0.

3. <u>Preparation of α-D-Desthiobiotin-ε-Cy5-L-lysine-MESNA [N^6-(6-{(2*Z*)-2-[(2*E*,4*E*)-5-(1-ethyl-3,3-dimethyl-5-sulfo-3*H*-indolium-2-yl)penta-2,4-dienylidene]-3,3-dimethyl-5-sulfo-2,3-dihydro-1*H*-indol-1-yl}hexanoyl)-N^2-(6-{[6-(5-methyl-2-oxoimidazolidin-4-yl)hexanoyl]amino}hexanoyl) lysylthio ethane-2-sulfonic acid]</u>



3.1 <u>Preparation of α-Fmoc-ε-Cy5-L-lysine-OH [2-[(1*E*,3*E*)-5-(1-{6-[(5-carboxy-5-{[(9*H*-fluoren-9-ylmethoxy)carbonyl]amino}pentyl)amino]-6-oxohexyl}-3,3-dimethyl-5-sulfo-1,3-dihydro-2*H*-indol-2-ylidene)-1,3-pentadienyl]-1-ethyl-3,3-dimethyl-5-sulfo-3*H*-indolium salt]</u>

Cy5 mono free acid potassium salt (Amersham Pharmacia Biotech Ltd) (450mg, 0.65 mmol) and DIEA (720μl) were dissolved in anhydrous dimethylsulphoxide (18ml). To this was added O-(N-succinimidyl)-N,N,N',N'-bis(tetramethylene)-uronium hexafluorophosphate (666mg, 1.6mmol) and the reaction mixture stirred at room temperature for 1 hr after which time negligible starting material remained by TLC (RPC_{18}, 1:1 methanol:water). The reaction mixture was slowly poured into diethyl ether to precipitate the product; Cy5 mono NHS ester, which was filtered off, washed with ethyl acetate and dried *in vacuo*. The product was re-dissolved in anhydrous

dimethylsulphoxide (18ml) and DIEA (720µl) added. Fmoc-lysine-OH (360mg, 0.98mmol) was suspended in a mixture of phosphate buffer (pH 7.4) (9ml) and dimethylsulphoxide (9ml). The suspension was slowly added to the solution of Cy5 NHS ester. The reaction mixture was stirred at room temperature for 12 hours. TLC (RPC_{18}, 2:3 methanol:water) showed the disappearance of starting material and the formation of a new product spot.

The product was purified by HPLC (Dynamax C18 column (50 x 4.14cm); flow rate 25ml/min; gradient of 20 to 80% B over 80 mins (eluent A = 0.1% TFA in water and eluent B = 0.1% TFA in acetonitrile); detection at 650nm. The fractions containing the desired product were pooled and most of the solvent removed under reduced pressure, the residue was freeze dried. The product; α-Fmoc-ε-Cy5-L-lysine-OH was obtained as a fluffy cyan solid (487mg, 74%). MS (MALDI TOF) found 1008(M^+); [theoretical ($C_{54}H_{64}N_4O_{11}S_2$) 1009]. ^1H NMR (200 MHz D_6DMSO) 1.27(t, 3H, CH_3), 1.35 (m, 4H, CH_2, CH_2), 1.55 (m's, 4H, CH_2, CH_2), 1.7 (s, 12H, $(CH_3)_2$), 1.78 (m, 2H, CH_2), 2.05 (t, 2H, CH_2), 3.0, (m, 2H, CH_2), 3.92 (m, 1H CH amino acid), 4.11 (m, 4H, N-CH_2, N^+CH_2), 4.27 (m 3H O-CH_2, CH fluorenyl), 6.3 (d, 2H, α, α' methine), 6.59, (t, 1H, γ methine) 7.28-7.48 (m's, 6H, Fmoc and indole Ar), 7.65 (d, 2H, fluorenyl Ar), 7.73 (d, 2H, fluorenyl Ar), 7.85 (s, 2H, indole Ar), 7.9 (d, 2H, indole Ar), 8.38 (t, 2H, β, β' methine).

3.2 <u>Preparation of ε-Cy5-L-lysine-OH [*N*⁶-(6-{(2*E*)-2-[(2*E*,4*E*)-5-(1-ethyl-3,3-dimethyl-5-sulfo-3*H*-indolium-2-yl)penta-2,4-dienylidene]-3,3-dimethyl-5-sulfo-2,3-dihydro-1*H*-indol-1-yl}hexanoyl)lysine]</u>

α-Fmoc-ε-Cy5-L-lysine-OH (100mg, 0.1mmol) was deprotected in a mixture of 20% piperidine in NMP (2ml). TLC (RP C18; 1:1 MeOH:water) showed the formation of a new product spot, rf = 0.92 as compared to that of the starting material, rf = 0.46. The piperidine was removed under reduced pressure and the dye precipitated by pouring the reaction mixture into diethyl ether. The product was filtered off and washed with dichloromethane and

then ethyl acetate to remove the yellow Fmoc derived by-product. The product was dissolved in water, filtered and then purified by HPLC; [Vydac Protein and Peptide C18 column; 0-50%B over 45mins at 10ml/min; eluent A = 0.1% TFA/water, eluent B = 0.1% TFA/MeCN, detection at 215nm). Fractions containing the desired product were combined and the solvents removed under reduced pressure to leave a blue residue. The residue was triturated with ethyl acetate and the resultant solid dried under vacuum at 40°C. The product; ε-Cy5-L-lysine-OH was obtained as a dark blue solid (43mg, 48%). Analytical HPLC AKTA analysis; Phenomenex C18 column; 0-50%B over 30mins at 1ml/min; eluent A = 0.1% TFA/water, eluent B = 0.1% TFA/MeCN, detection at 650nm; rt = 20.22mins. MS (MALDI TOF) found 785 (M^+); [theoretical ($C_{39}H_{53}N_4O_9S_2$) 785]. ^1H NMR (300 MHz D_6DMSO) 1.26 (t, 3H, CH_3), 1.52 (m, 4H, CH_2, CH_2), 1.62 (m, 4H, CH_2, CH_2), 1.69 (S, 12H, $(CH_3)_2$), 2.02 (m, 2H, CH_2), 2.92 (m, 2H, CH_2), 3.85 (m, 1H, CH amino acid), 4.10 (m, NCH_2, N^+CH_2), 6.29 (d, 1H α methine), 6.34 (d, 1H, α' methine), 6.58 (t, 1H, γ methine), 7.32 (m, 2H, indole Ar), 7.82 (d, 2H, indole Ar), 8.04 (m, 3H, NH_3^+), 8.37 (t, 2H, β, β' methine).

3.3 <u>Preparation of D-Desthiobiotinamidocaproic acid</u>

D-Desthiobiotin (250mg, 1.17mmol) was dissoved in anhydrous dimethylsulphoxide (2ml). To this solution was added PyAOP (610mg, 1.17mmol) and DIEA (200μl, 1.15mmol). The mixture was stirred under nitrogen at RT for 3hrs before adding 6-aminocaproic acid (153mg, 1.17mmol) and a further amount of DIEA (200μl, 1.15mmol). The reaction mixture was stirred for a further 4hrs. TLC (RP C18; 1:2 MeOH:water; detection by cinnamaldehyde staining) showed the formation of a new product spot, rf = 0.63 as compared to the starting material, rf = 0.76). The reaction mixture was poured into excess diethyl ether to give a brown oil. The oil was triturated with ethyl acetate until an off-white solid was obtained. The product was filtered off and purified by HPLC [Vydac Protein and Peptide C18 column; 0-50%B over 30mins at 10ml/min; eluent A = 0.1% TFA/water, eluent B = 0.1% TFA/MeCN, detection at 215nm). Fractions containing the desired product

were pooled and the solvents removed under reduced pressure. The residue was triturated with ethyl acetate to give a white solid. The product was filtered off and dried under reduced pressure at 40°C. The product; D-desthiobiotinamidocaproic acid, was obtained as a white solid (48mg, 12.5%). MS (MALDI TOF) found 327(M^+); [theoretical ($C_{16}H_{29}N_3O_4$) 327]. ^1H NMR (300 MHz D_6DMSO) 0.96 (d, 3H, CH_3), 1.25 (m, 6H, CH_2, CH_2, CH_2), 1.34 (m, 4H, CH_2, CH_2), 1.48 (m, 4H, CH_2, CH_2), 2.03 (m, 2H, C(O)CH_2), 2.60 (m, 2H, C(O)CH_2), 3.01 (m, 2H, NHCH_2), 3.47 (m, 1H, CH), 3.60 (m, 1H, CH), 6.11 (s, 1H, NH), 6.29 (s, 1H, NH), 7.71 (s, 1H, NH).

3.4 <u>Preparation of D-Desthiobiotinamidocaproate N-hydroxy succinimidyl ester</u>

D-Desthiobiotinamidocaproic acid (48mg, 0.147mmol) was dissolved in DMF (1ml) and N,N,N',N'-bis(tetramethylene)-O-(N-succinimidyl)uronium hexafluorophosphate (HSPyU) (90mg, 0.21mmol) and DIEA (40µl, 0.23mmol) were added. The reaction mixture was stirred under nitrogen at RT for 6hrs, TLC (RP C18; 1:2 MeOH:water; materials detected by cinnamaldehyde staining) showed the formation of a new product at the base line as compared to the starting material, rf = 0.68. The reaction mixture was poured into diethyl ether to give a brown gum. The supernatant was decanted off and the gum again treated with diethyl ether. No solid formed. The gum was dried under reduced pressure and the product, D-desthiobiotinamidocaproate N-hydroxy succinimidyl ester was used directly in the next dye coupling reaction, assuming a theoretical yield of 62mg.

3.5 <u>Preparation of α-D-Desthiobiotin-ε-Cy5-L-lysine-OH [*N*^{6}-(6-{(2*Z*)-2-[(2*E*,4*E*)-5-(1-ethyl-3,3-dimethyl-5-sulfo-3*H*-indolium-2-yl)penta-2,4-dienylidene]-3,3-dimethyl-5-sulfo-2,3-dihydro-1*H*-indol-1-yl}hexanoyl)-*N*^{2}-(6-{[6-(5-methyl-2-oxoimidazolidin-4-yl)hexanoyl]amino}hexanoyl)lysine]</u>

ε-Cy5-L-lysine-OH (43mg, 0.048mmol), D-desthiobiotinamidocaproate N-hydroxy succinimidyl ester (62mg, 0.146mmol) and DIEA (80µl, 0.45mmol)

were stirred together in DMF (2ml) for 3hrs. TLC (RP C18; 1:1 MeOH:water) showed the formation of a new product spot, rf = 0.79, just under that of the starting material. The product was precipitated into diethyl ether (200ml) and then filtered off. The material was purified in multiple runs by HPLC [Vydac Protein and Peptide C18 column; eluent A = 0.1% TFA/water, eluent B = 0.1% TFA/MeCN, various gradients, detection at 215nm) until the material was seen to be pure by ^{1}H NMR. Analytical HPLC AKTA analysis; Phenomenex C18 column; 0-50%B over 30mins at 1ml/min; eluent A = 0.1% TFA/water, eluent B = 0.1% TFA/MeCN, detection at 650nm; rt = 22.04mins. MS (MALDI TOF) found 1094 (M^+); [theoretical ($C_{55}H_{80}N_7O_{12}S_2$) 1094].

3.6 Preparation of α-D-Desthiobiotin-ε-Cy5-L-lysine-MESNA [N^6-(6-{(2*Z*)-2-[(2*E*,4*E*)-5-(1-ethyl-3,3-dimethyl-5-sulfo-3*H*-indolium-2-yl)penta-2,4-dienylidene]-3,3-dimethyl-5-sulfo-2,3-dihydro-1*H*-indol-1-yl}hexanoyl)-N^2-(6-{[6-(5-methyl-2-oxoimidazolidin-4-yl)hexanoyl]amino}hexanoyl) lysylthio propane-3-sulfonic acid]

α-D-Desthiobiotin-ε-Cy5-L-lysine-OH (10mg, 8.8μmol) was dissolved in anhydrous dimethylsulphoxide (2ml), PyAOP (10mg, 19.2μmol), MESNA (5mg, 0.30mmol) and DIEA (10μl, 0.06mmol) were added and the reaction mixture was stirred under nitrogen for 4hrs. The reaction mixture was purified by RP-HPLC; [Vydac Protein and Peptide C18 column; 15-40%B over 45mins at 10ml/min; eluent A = 0.1% TFA/water, eluent B = 0.1% TFA/MeCN, detection at 215nm). The product containing fractions were combined and the majority of solvent removed under reduced pressure, the residue was freeze dried. The product was obtained as a fluffy blue solid (4mg, 37%). TLC; RP C18; 1:1 water:acetonitrile) rf = 0.76. Analytical HPLC AKTA analysis; Phenomenex C18 column; 0-50%B over 30mins at 1ml/min; eluent A = 0.1% TFA/water, eluent B = 0.1% TFA/MeCN, detection at 650nm; rt = 21.04mins. λmax = 648nm (PBS buffer). MS (MALDI TOF) found 1219 (MH^+); [theoretical ($C_{57}H_{84}N_7O_{14}S_4$) 1218].

4. Protein labelling and affinity purification

4.1 Labelling of N terminal cysteine Grb2SH2 with α-D-desthiobiotin-ε-Cy5-L-lysine-MESNA and purification

To N-terminal cysteine Grb2SH2 (N-Cys-Grb2SH2) (200μM in PBS buffer pH 7.2) (200μl) was added α-D-desthiobiotin-ε-Cy5-L-lysine-MESNA (2mM in reaction buffer) (200μl). N-Cys-Grb2SH2 was prepared using recombinant techniques. The reaction buffer consisted of phosphate buffer (200mM), pH 7.2 containing sodium chloride (200mM) and 4% MESNA. The reaction mixture was left at RT for 12 hrs, wrapped in foil to protect from light. The reaction was then quenched with dl-dithiothreitol (final concentration 60mM). Unreacted dye was separated from labelled/unlabelled protein by FPLC, using a fast desalt column and eluent of PBS buffer, pH 7.4; 2ml/min, detection 280 nm. Protein fractions were combined and desthiobiotin-ε-Cy5-L-lysine affinity probe labelled protein was bound to streptavidin beads (PIERCE Ultralink™ streptavidin). The beads were washed vigorously with both PBS buffer and binding buffer (PBS containing 500mM NaCl). The product; α-D-desthiobiotin-ε-Cy5-L-Lys-Cys-Grb2SH2 was extracted from the streptavidin beads by adding cold biotin (1.6mM). Several extraction runs were required. The materials were further purified by dialysis (PIERCE Slide-a-lyser™ mini dialysis units, 7,000 mwco) to remove free biotin from the sample. The product was analysed by SDS PAGE together with the following controls (see Figure1):

Lane 1	MW marker
Lane 2	Unligated protein control
Lane 3	Ligation reaction mixture: α-D-desthiobiotin-ε-Cy5-L-Lysine-MESNA and N-Cys-Grb2SH2
Lane 4	Labelled/unlabelled N-Cys-Grb2SH2 after FPLC purification
Lane 5	Unbound protein
Lane 6	Streptavidin bead washes
Lanes 7,8,9	Affinity purified product
Lane 10	Unreacted α-D-desthiobiotin-ε-Cy5-L-Lysine-MESNA

The gel was imaged using a Typhoon imager (Figure 1B) using parameters for Cy5 fluorescence to detect fractions containing the fluorescent label. The gel was then stained with Coomassie blue stain (Figure 1A) to determine the protein containing fractions. SDS PAGE gel shows that (a) unlabelled protein (both factor XA and N-Cys-Grb2SH2 did not bind to the streptavidin beads (Figure 1A and 1B, column 5) (enriched protein stain) and (b) the product was removed from the streptavidin beads by adding cold biotin (Figure 1A and 1B, columns 7 and 8) (both protein stain and Cy5 fluorescence).